

WHAT IS CLAIMED IS:

Sub A³ 7 1. A method for assigning traffic buckets to a cache system, the method comprising:

5 when a new cache system starts up in a cache cluster having a plurality of total buckets, determining a full bucket allocation for the new cache system;

assigning buckets to the new cache system using a first technique when the cache cluster is not operating at a maximum load; and

10 assigning buckets to the new cache system using a second technique that differs from the first technique, wherein the second technique is performed after the first technique.

000E90" 64580960
15 2. A method as recited in claim 1 further comprising assigning the full bucket allocation to the new cache system when the cache cluster is operating at a maximum load.

20 3. A method as recited in claim 1 further comprising assigning buckets to the new cache system based on a first technique when the cache cluster is not operating at a maximum load.

4. A method as recited in claim 1, the first technique comprising:
periodically monitoring a load of the new cache system

5u A³ 7

when the new cache system is overloaded, shedding a minimum number of buckets from the new cache system; and

when the new cache system is underloaded, adding the minimum number of buckets to the new cache system.

5

5. A method as recited in claim 4 further comprising:

periodically monitoring a load of each of the other cache systems within the cache cluster;

when any of the other cache system is overloaded, shedding the minimum number of buckets from the overloaded other cache system; and

when any of the other cache system is underloaded, adding the minimum number of buckets to the underloaded other cache system.

6. A method as recited in claim 5, wherein the minimum number of buckets equals a single bucket.

7. A method as recited in claim 4 wherein the second technique is performed until the full allocation has been assigned to the new cache system or a minimum number of buckets have been added to or shed from the new cache system, the second technique comprising:

initially assigning a portion of the full bucket allocation to the new cache system;

when the new cache system is overloaded and when buckets have been previously shed, periodically shedding a portion of a number of buckets that were previously shed from the new cache system;

when the new cache system is not overloaded and when buckets have been previously shed, periodically assigning a portion of a number of buckets that were previously shed from the new cache system.

9. A method as recited in claim 7 wherein the portion of the number of buckets that were previously shed from the new cache system, the portion of the unassigned buckets, and the portion of the assigned buckets are equal to a half portion.

29

Sub A³ 7

when an existing cache system leaves the cache cluster or shuts down,
determining a new bucket allocation for each of the remaining cache systems;
and

assigning buckets to the remaining cache system using the first
technique.

11. A method as recited in claim 7 further comprising:

when an existing cache system leaves the cache cluster or shuts down,
determining a new bucket allocation for each of the remaining cache systems;
and

assigning buckets to the remaining cache system using the first
technique.

12. A method as recited in claim 1, wherein the full bucket allocation is equal to a
number of buckets allocated to each existing cache system within the cache cluster.

13. A method as recited in claim 1, wherein the full bucket allocation is not equal
to a number of buckets allocated to each existing cache system within the cache cluster.

14. A method as recited in claim 13, further comprising receiving a weight value
from the new cache system indicating a percentage of the total buckets to allocate the new
cache system.

Sub A³ 7

21. A computer system as recited in claim 20, wherein the second technique is performed until the full allocation has been assigned to the new cache system or a minimum number of buckets have been added to or shed from the new cache system, the second technique comprising:

5 initially assigning a portion of the full bucket allocation to the new cache system;

when the new cache system is overloaded and when no buckets have been previously shed, periodically shedding a portion of the assigned buckets from the new cache system;

10 when the new cache system is overloaded and when buckets have been previously shed, periodically shedding a portion of a number of buckets that were previously shed from the new cache system;

15 when the new cache system is not overloaded and when no buckets have been previously shed, periodically assigning a portion of the unassigned buckets to the new cache system; and

when the new cache system is not overloaded and when buckets have been previously shed, periodically assigning a portion of a number of buckets that were previously shed from the new cache system.

20 22. A computer system as recited in claim 21 wherein the portion of the number of buckets that were previously shed from the new cache system, the portion of the unassigned buckets, and the portion of the assigned buckets are equal to a half portion.

Sub A³ 7

23. A computer system as recited in claim 21 wherein at least one of the memory and the processor are further adapted to provide:

when an existing cache system leaves the cache cluster or shuts down,
determining a new bucket allocation for each of the remaining cache systems;
and
assigning buckets to the remaining cache system using the first technique.

24. A computer system as recited in claim 18 wherein at least one of the memory and the processor are further adapted to provide:

when an existing cache system leaves the cache cluster or shuts down,
determining a new bucket allocation for each of the remaining cache systems;
and
assigning buckets to the remaining cache system using the first technique.

25. A computer program product for assigning traffic buckets to a cache system, the computer program product comprising:

at least one computer readable medium;
computer program instructions stored within the at least one computer readable product configured to cause a processing device to:
when a new cache system starts up in a cache cluster having a plurality of total buckets, determine a full bucket allocation for the new cache system;

